DOCUMENT RESUME

ED 273 188 HE 019 601

TITLE	Refinements to NCHEMS Data Base Software and Holdings, FY82.
INSTITUTION	National Center for Higher Education Management Systems, Boulder, Colo.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	[83]
NOTE	15p.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Computer Software; *Databases; *Data Collection; *Higher Education; *Information Centers; *National Surveys
IDENTIFIERS	*National Center for Higher Educ Management Systems

ABSTRACT

Holdings of the National Center for Higher Education Management Systems (NCHEMS) and fiscal year (FY) 1982 data acquisitions and software enhancements are discussed. NCHEMS holds many years of data from surveys conducted by the National Center for Education Statistics, the National Science Foundation, the Bureau of the Census, and the Department of Commerce. NCHEMS has also built a base of standardized software to increase the productive capability of the NCHEMS staff. Enhancements to the standard software available at NCHEMS have been accomplished through: in-house development of new programs, addition of enhancements to in-house developed programs, and installing or enhancing externally-purchased programs. The most critical of new programs developed in-house is used to calculate and assign institutional classification according to the system developed at NCHEMS in the late 1970s. Frequent demand for counts of degree programs of different sorts led to the development of a general purpose program to generate a master file containing counts at the four-digit level for all bachelor, master, and doctoral programs recorded in the Higher Education General Information Survey earned degrees survey. Data sets that were acquired in FY 1982 are listed. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED273188

HE

PC-2A

Refinements to NCHEMS

Data Base Software and Holdings

FY82

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

NCHEMS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

NE 01 9 601

National Center for Higher Education Management Systems
P.O. Drawer P
Boulder, CO 80302

Refinements to NCHEMS

Data Base Software and Holdings

FY82

Introduction

Data concerning higher education in the United States is collected by a variety of governmental agencies, higher-education organizations, and individual researchers. The diversity of collection purposes, timeframes, data definitions, and respondents leads to complex problems in converting the data to formats appropriate to uses different from those of the original data collectors. In order to address this problem, The National Center for Higher Education Management Systems (NCHEMS) developed and maintains the most comprehensive data base about higher education in the country. Such a centralized repository where diverse sets of data are compiled and where expertise is concentrated contributes to the efficiency and efficacy of data use. It also provides access to a wide variety of users who otherwise would be denied access by virtue of their lack of the resources needed to perform their own data acquisition and conversion. Maintaining the currency and relevance of the data base holdings and enhancing the software needed for smooth, efficient access to these holdings is thus an on-going activity at NCHEMS.

Data Base Holdings

In order to keep NCHEMS data holdings current, data were acquired during FY82 from a number of sources and added to or used to update the NCHEMS data bases. The primary sources of data have been the National Center for Education Statistics (NCES), the National Science Foundation (NSF), the Bureau of Census, and the Department of Commerce. A list of the data added is given in Table A.

Table A

Additions to Database Holdings

Description	Source
HEGIS 15 Institutional Characteristics 1980-81	NCES
HEGIS 15 Opening Fall Enrollment 1980	NCES
HEGIS 15 Financial Statistics FY80	NCES
HEGIS 15 Earned Degrees and Other Formal Awards Conferred 1979-80	NCES
HEGIS 15 Salaries and Benefits of Full-Time Instructional Faculty 1980-81	NCES
Survey of Scientific and Engineering Employment at Universities and Colleges January 1981	NSF
Survey of Graduate Science Students and Postdoctorals Fall 1980	NSF
Survey of Scientific and Engineering Expenditures at Universities and Colleges FY80	NSF
High School and Beyond Student File	NCES
High School and Beyond School File	NCES
High School and Beyond Parent File	NCES
High School and Beyond Language File	NCES
Personal Income Per Capita 1979	Dept. of Commerce
State and Local Revenues Per Capita 1978-79	Dept. of Commerce
State and Local Expenditures 1978-79	Dept. of Commerce
Tax Capacity Index and Tax Effort Index 1979	Halstead
State Scholarships 1978-79	SHEEO/NCES Network
Total Population 1978, 1979	Bureau of Census
Population breakdown by age 1973, 1978, 1979	Bureau of Census

Table A
(continued)

Additions to Database Holdings

Description	Source
HEGIS 16 Opening Fall Enrollment 1981	NCES
HEGIS 16 Institutional Characteristics 1981-82	NCES
HEGIS 16 Financial Statistics FY81	NCES
Classification of Instructional Programs Crosswalk	NCES
Selected data from Governmental Finances in 1979-80	Dept. of Commerce
Selected data from Voluntary Support of Education 1979-80	Council for Financial Aid to Education
Selected data from Economic Indicators	Council of Economic Advisors

The new data were used to produce updated versions of both the long and short formats of our standard in-house HEGIS master files for fall enrollment, finance, earned degrees, and faculty salaries. In-house master files of NSF student, expenditure, and employees data and the statewide analysis file were also brought up-to-date with changes from the newly acquired files. Tapes containing data from the NCES surveys referred to as High School and Beyond (student, school, parent, and language), were tested to make sure that they conformed to the documentation. Since the researchers originally requesting these data have not yet actually needed them, no master subset has yet been specified, delaying the process of formatting and building standard master files.

Multiple demands for more detailed HEGIS data than that available in even the long formats of standard master files prompted the design and creation of new master file formats for opening fall enrollment and earned degrees. The new master file for fall enrollment contains ethnic data and totals for all levels of detail collected on the summary (9999) page of the HEGIS questionnaire. The new earned degrees file contains only the ethnic data collected by the HEGIS questionnaire. Detail enrollment masters were built for 1976-77, 1977-78, 1978-79, 1979-80, and 1980-81. Earned degrees ethnic-data-only masters were built for 1975-76, 1976-77, and 1978-79. Use of these new master files rather than the original transaction files created by NCES simplifies access to detailed ethnic data and substantially reduces the number of records processed when such data are needed. A new master format was also developed for information from the HEGIS directory file, since we had repetitive requests for items that are not included in any of our other standard master file formats. The new directory files again take the place of the complex processing required to obtain a few necessary items from the original transaction tapes. New directory masters were built for the years 1972-73 through 1980-81 inclusive.

A succession of requests for a variety of counts of fields of study at the four-digit level of the HEGIS taxonomy of instructional programs in which degrees were granted led to the development of a count master file. The count master file contains counts by field of the number of sub-fields in which an institution granted bachelor, master, doctorate degrees or any of the three. This master allows calculation of any of the degree counts requested to date without generating an elaborate custom count program. The file was built only for the most recent data, 1979-80, but can easily be generated for other years if the need arises (see discussion of software enhancements).

Another new master file which has proven exceptionally useful is the preliminary version of the universe file developed for the HEGIS longitudinal study. Since any trend analysis using HEGIS data must allow for changes in the universe, this file has seen a great deal of in-house use. Because of its relevance to so much of our in-house work, we will continue to update the universe file internally as new data become available.

Software Refinements

Data, in themselves, are not very useful. The recognition of this is reflected in the evolution of jargon in the computer field from "data processing" to "information processing." In order to convert "data" into "information" a wide variety of programs are needed to separate, combine, compare and otherwise manipulate the individual items referred to as data. Each analysis can be regarded as an independent entity requiring its own custom-tailored data files as well as a set of unique programs to create and manipulate those data files. However, such an approach is labor and time intensive although it tends to use the actual computer resources more efficiently. As the cost of computer equipment has fallen, the cost of human resources needed for programming has

become more critical. Although there is continuing need for some amount of custom programming in any computer application, quite a few tasks are common to many applications. These common tasks can be streamlined so that they use much less of the expensive, scarce human resources. Use of general purpose software packages which contain routines that perform functions that otherwise are tricky and/or time consuming to program, requires comparatively simple specifications to custom-tailor to a particular application.

NCHEMS has made a concerted effort to build a base of standardized software in order to increase the productive capability of the NCHEMS staff. After approximately three years, NCHEMS currently has software to facilitate subsetting, matching, updating, and aggregating data when building data files, selecting and formatting data when generating tables and reports, and combining, comparing and manipulating data for statistical analysis. In additon, several series of programs that perform similar functions, which had been developed in a custom-tailored manner to handle variations in data files or researcher needs, have been replaced by single, broader-based programs that provide more continuity and more assurance of comparability. Enhancements to the standard software available at NCHEMS have been accomplished in four ways, through:

- in-house development of new programs
- addition of enhancements to in-house developed programs
- installation of enhancements to externally purchased programs
- conversion/installation of externally developed programs.

The following discussion provides a brief description of the software development and enhancement activities undertaken during FY82.

The most critical of the new programs developed in-house is used to calculate and assign institutional classification according to the system developed at NCHEMS in the late 1970's. Previous work on NCHEMS classification was performed via a service bureau on an IBM mainframe. In order to reduce expenses and to assure compatability with the data used in NCHEMS other work, these programs were redesigned and rewritten to run in-house on the HP-3000. The main program, written in FORTRAN, uses HEGIS earned degree and finance data to calculate NCHEMS classification following an algorithm based on the one illustrated in "An Improved Taxonomy of Postsecondary Institutions" by David Makowski and Rolf M. Wulfsberg, March 1980. The calculated classification is then compared to a history file containing the prior years' calculated and assigned classifications and the new classification is automatically assigned. To test the program and assure historical continuity, the program was used to build a new history file beginning with 1972 data. In addition, the HEGIS universe file was used to assign inapplicable codes where appropriate to units that opened or closed during the period covered and to those units, such as system offices, that are not HEGIS reporting units. We are now assured that when these classification codes are used in an analysis or report that they are consistent with the data in the standard master files and that there is historical continuity in the calculation. Codes based on new data can also be readily calculated in a consistent manner. Modifications to the master files and to the main classification program will need to be made in the future when NCES begins collecting earned degree data using the new classification of instructional programs.

Prior to purchasing the HP-3000, much of NCHEMS work utilized a file management package named MARKIV. In order to facilitate staff transition to the new machine, some of the early software developed utilized input and output

formats similar to those used in MARKIV. However, to provide a more efficient means to generate file documentation as well as a format that is more easily read than the MARK IV glossary format, new file documentation software was developed. The program, written in FORTRAN, uses a simple "label file" to access our standard in-house file definitions and generate documentation that indicates the starting and ending positions, name, length, and description of each variable in the file. An automated checking feature may be invoked to flag discontinuities in the description of the file and guard against errors in file building resulting from overlapping or discontinuous fields. Since the program allows for longitudinal data, minor modifications will need to be made annually to incorporate newly released data.

Frequent demand for counts of degree programs of different sorts led to development of a general purpose program to generate a master file containing counts at the four-digit level for all bachelor, master, and doctoral programs recorded in the HEGIS earned degrees survey. In addition, several programs were previously necessary to calculate program counts at the two-digit level due to minor historical changes in the survey. These were replaced with a single standard program that functions for all years to date. The two programs replace approximately ten programs developed to handle special purpose counts needed in the past. Both programs were written in FORTRAN and will need to be modified in the future to accomodate the new NCES classification of instructional programs.

The development of standard master file formats for detailed ethnic data for earned degrees and fall enrollment also made it necessary to produce standard programs and procedures for accessing these files. Standard definitions and protocols have been developed. However, they still need to be incorporated in our standard in-house database documentation which is currently undergoing

revision to include a variety of updates and changes that have occurred since the documentation format was implemented. In addition revisions were made to the procedures to build the standard earned degree masters in order to incorporate the two-digit program counts generated by the new standard program described previously. Revisions were also necessary to the programs that build the standard employees masters. In the past NCHEMS has built its master files by using an in-house suppression program on the unsuppressed version of NCES data files. The in-house version suppressed all salary cells for subsets containing three or fewer individuals. Due to changes in staff, the unsuppressed tapes were unavailable and the NCHEMS programs had to be converted to work on the public-use data files which have already been suppressed by NCES. NCES suppresses cells containing two or fewer individuals, so the master files now contain some historical discontinuity.

One of the most heavily used of the general purpose software systems developed in-house is the transaction processor which is used to build or update data files. Necessary input to the transaction processor is developed through programs used to create file definitions (computable file descriptions supplying name and location of every data element in the file) and transaction definitions (computable descriptions of the operations to be performed during transaction processing, e.g., move whatever is found at the specified location to the specified data element in a new file). A number of additions were made to these programs to further increase their utility. The program used to interactively enter file definitions was enhanced to include an option whereby the program calculates and automatically enters the start location of the next sequential variable. This enhancement reduces the time needed to enter file definitions and prevents the offset calculation errors that were possible when a programmer had to hand calculate and enter the start location. The capability to exit from the

program and then resume work on the same definition at a later time and the capacity to display the last entered variable and current location were also added. Both were needed to avoid the complications of creating and combining the partial definitions that resulted if the entry process was interrupted. Similar enhancements to allow exit and reentry and to display the last entry were added to the program used to interactively enter transaction definitions.

Several enhancements were also added to the transaction processor itself. The most important of these was the capability to create a new master in addition to retaining the old master. This obviated the need to back up the old master prior to running an update and provided the ability to add new records to an extant file. An audit option was also added that allows the listing of rejected transactions in order to trace problems in a particular transaction update run.

The other in-house developed system that underwent substantial change was the procedure used to convert packed data to decimal data for use with statistical software. This procedure was modified to improve efficiency. Fewer steps, less disc space, and less time is now required to convert large packed data files like some of the early HEGIS transaction files created by NCES.

Software in general is constantly undergoing change in order to keep abreast with advancements in technology. The most critical software system on general purpose computers is the operating system which handles all communications between users and the hardware. When a computer vendor makes changes to the operating system, it also changes any other software it markets to conform to the new version of the operating system. Other software manufacturers must also sometimes modify their programs because of changes in the operating system. In order to assure that our software systems remain functional, NCHEMS retains software maintenance contracts on all externally purchased software. The

maintenance contracts entitle NCHEMS to any updates the software vendor makes to the product, either to conform to changes in the operating system or to improve the package's capabilities.

A variety of enhancements to externally purchased software were tested and installed during FY82. A large number of new capabilities were added with installation of a new version of the report generation software, QUIZ. Among the most important changes were new options for producing subfiles--including improved syntax for subfiling, sorting, and generating control breaks--and the ability to save compiled programs. The former, in addition to reducing necessary programmer effort, has provided a means for avoiding a particularly nasty operating system bug that, on occasion, brought the whole computer down. The latter allows for more efficient running of standard reports. In addition, eight new functions, including variable subscripts, were added to the capability to define temporary variables. Five new run-time parameters, the capacity to check optional co-ordinated files, and a capacity for parallel as well as hierarchical file structures are other useful new features.

Enhancements to the statistical package, SPSS, included a box jenkins routine to fit and forecast time series data and a new regression routine. Additional edit capabilities were added with the installation of SPSS EDIT, a system of routines that allow use from within SPSS of external editing software. We are still waiting for correction of bugs in the routines that allow use of tape storage rather than disc for SPSS files.

Enhancements to the word processing/editor software, TDP, were delayed when testing revealed severe problems with the new software and it was decided to return to the old version until the vendor had repaired the bugs found in the new version.

Several programs developed by other users of HP-3000 systems were also installed during the last year. Access to these programs is obtained by participation in computer users groups and in classes and seminars at Hewlett-Packard. Improved documentation for the chart files used in generating graphs, capability to list creators, along with access and update dates for disc files, and a set of utilities that allow reporting of information collected in the computer log files are among the additions that improve NCHEMS ability to monitor computer usage and help users control their use of the system and reduce their costs.

Beyond the enhancements to the data base and software, a major contribution to using NCHEMS resources was added with the in-house publication and distribution of the first edition of "The NCHEMS/WICHE Guide to Using the HP-3000." In addition to describing the facilities and schedules, this 75 page document provides instructions for running general purpose software developed in-house and brief descriptions of externally developed software along with references to the detailed documentation available for each. A list of the major subsections of this document is given in Table B. A copy of the entire publication is enclosed for the information of those interested in greater detail.

TABLE B

NCHEMS/WICHE Guide to Using the HP3000

Introduction
Section A

1. Schedules
2. Operations
3. Facilities
4. Getting Started
5. Hardware
6. Software
7. Tape Library
8. Data Entry
9. Training
10. Technical Consulting
11. Programming Assistance
12. Chargebacks
13. Refunds
14. Hardware Acquisition

Section B

1.	TDP	(word processing)
2.	QUIZ	(report generation)
3.	SPSS	(statistics)
4.	BMDP	(statistics)
5.	KBS	(bibliography)
6.	DSG/3000	(graphics)
7.	PLOT/21	(graphics)
8.	MAP	(maps)
9.	MAILER	(mailing labels)
10.	TRANSACT	(file building)
11.	SUBFILER	(file building)
12.	COORDPRG	(file building)
13.	MCOPY2	(copy utility)
14.	EXTRACT	(file building)
15.	CBLOCK	(disc utility)
16.1	MARKIV	(file definition)
16.2	TRANDEF	(transaction definition)
16.3	CNVTUDC	(file definition)
16.4	SPSSDEF	(file definition)
16.5	GLOSS	(file definition documentation)
17.1	BLOCK	(disc utility)
17.2	TAPEMAP	(tape utility)
17.3	TAPESCAN	(tape utility)
17.4	SCRUNCH	(disc utility)
17.5	PSCREEN	(terminal utility)
17.6	SLEEPER	(delayed job submission utility)
18.1	STORE	(back-up utility)
18.2	VALIDATE	(back-up utility)
18.3	RESTORE	(back-up utility)